Journal of Educational Psychology
2015, Vol. 107, No. 1, 258–271

0022-0663/15/$12.00 http://dx.doi.org/10.1037/a0037485

The Big-Fish-Little-Pond Effect: Generalizability of Social Comparison Processes Over Two Age Cohorts From Western, Asian, and Middle Eastern Islamic Countries

Herbert W. Marsh
Australian Catholic University, King Saud University, and University of Oxford

Adel Salah Abduljabbar
King Saud University

Alexandre J. S. Morin and Philip Parker
Australian Catholic University

Faisal Abdelfattah
King Saud University

Benjamin Nagengast
University of Tübingen

Maher M. Abu-Hilal
Sultan Qaboos University

Extensive support for the seemingly paradoxical negative effects of school- and class-average achievement on academic self-concept (ASC)—the big-fish-little-pond effect (BFLPE)—is based largely on secondary students in Western countries or on cross-cultural Program for International Student Assessment studies. There is little research testing the generalizability of this frame of reference effect based on social comparison theory to primary school students and or to matched samples of primary and secondary students from different countries. Using multigroup–multilevel latent variable models, we found support for developmental and cross-cultural generalizability of the BFLPE based on Trends in International Mathematics and Science Study data; positive effects of individual student achievement and the negative effects of class-average achievement on ASC were significant for each of the 26 groups (nationally representative samples of 4th- and 8th-grade students from 13 diverse countries; 117,321 students from 6,499 classes).

Keywords: big-fish-little-pond effect, Trends in International Mathematics and Science Study, frame of reference effects, contextual effects, doubly-latent multilevel models

Supplemental materials: http://dx.doi.org/10.1037/a0037485.supp

Self-concept, dating back to at least the seminal work by William James (1890/1963), is one of the oldest and most important constructs in the social sciences. Today positive self-beliefs are also at the heart of a positive revolution sweeping psychology, which emphasizes focusing on how healthy, normal, and exceptional individuals can get the most from life (e.g., Bandura, 2006; Bruner, 1996; Diener, 2000; Marsh & Craven, 2006; Seligman & Csikszentmihalyi, 2000). Thus, self-concept enhancement is now a major goal in many fields including education, child development, health, sport/exercise sciences, social services, and management (Marsh, 2007). Self-concept is also an important mediating factor that facilitates the attainment of other desirable outcomes. Particularly in education settings, a positive academic self-concept (ASC) is both a highly desirable goal and a means of facilitating subsequent learning and other academic accomplishments. Our study is based on the 2007 Trends in International Mathematics and Science Study (TIMSS), with nationally representative samples from different countries and age cohorts, to provide tests of the developmental and cross-sectional generalizability of strong theoretical models of self-concept using new and evolving statistical methodology.

Big-Fish-Little-Pond Effect (BFLPE): The Theoretical and Substantive Focus

BFLPE Theoretical Models

Self-concept theory emphasizes that perceptions of the self cannot be adequately understood if frames of reference are ignored

This article was published Online First September 15, 2014.

Herbert W. Marsh, Institute for Positive Psychology and Education, Australian Catholic University; Department of Education, King Saud University; and Department of Education, University of Oxford; Adel Salah Abduljabbar, Department of Education, King Saud University; Alexandre J. S. Morin and Philip Parker, Institute for Positive Psychology and Education, Australian Catholic University; Faisal Abdelfattah, Department of Education, King Saud University; Benjamin Nagengast, Center for Educational Science and Psychology, University of Tübingen; Maher M. Abu-Hilal, Department of Psychology, Sultan Qaboos University.

This article was made possible in part by a grant from the Australian Research Council (DP130102713). We would like to thank Tihomir Asparouhov, Matthias Von Davier, Anna Preuschoff, and Michael Martin for helpful comments at earlier stages of this research.

Correspondence concerning this article should be addressed to Herbert W. Marsh, Institute for Positive Psychology and Education, Australian Catholic University, Locked Bag 2002, Strathfield NSW 2135, Australia. E-mail: herb.marsh@acu.edu.au

(Marsh, 2007). The same objective characteristics and accomplishments can lead to disparate self-concepts, depending on the frames of reference or standards of comparison that individuals use to evaluate themselves, and these self-beliefs have important implications for future choices, performance, and behaviors. From the time of William James (1890/1983), psychologists have recognized that objective accomplishments are evaluated in relation to frames of reference. Thus, James indicated, "we have the paradox of a man shamed to death because he is only the second pugilist or the second oarsman in the world" (p. 310). Marsh (1984; see also Marsh & Parker, 1984; Marsh, Seaton, et al., 2008) proposed the BFLPE to encapsulate frame of reference effects that are based on an integration of theoretical models and empirical research from diverse disciplines (e.g., relative deprivation theory, social comparison theory, psychophysical judgment, social judgment; see supplemental materials).

In the BFLPE model, students are hypothesized to compare their abilities with the abilities of their classmates and use this social comparison impression as one basis for forming their own self-concept (see Figure 1); individual ability is positively related to ASC (the brighter I am, the higher my ASC), but that class- and school-average ability have a negative effect on ASC (the brighter my classmates, the lower my ASC). Hence, ASC depends not only on a student's academic accomplishments but also on those of the student's classmates. Consistent with theoretical predictions and an increasing emphasis on the multidimensionality of self-concept, the BFLPE in academic settings is specific to ASC; class- and school-average ability achievement has little positive or negative effect on nonacademic components of self-concept or on global self-esteem (e.g., Marsh, 1987; Marsh, Chessor, Craven, & Roche, 1995; Marsh & Parker, 1984; for a review, see Marsh, Seaton, et al., 2008).

Diener and Fujita (1997, p. 350) reviewed BFLPE research in relation to the broader social comparison theory and concluded that BFLPE research provided the clearest support for predictions based on social comparison theory in an imposed social comparison paradigm. The reason for this, they surmised, was that the frame of reference, based on classmates within the same class or school, is more clearly defined in BFLPE research than in most other research settings. The importance of the class or school setting is that the relevance of the social comparisons in class or school settings is much more ecologically valid than manipulations

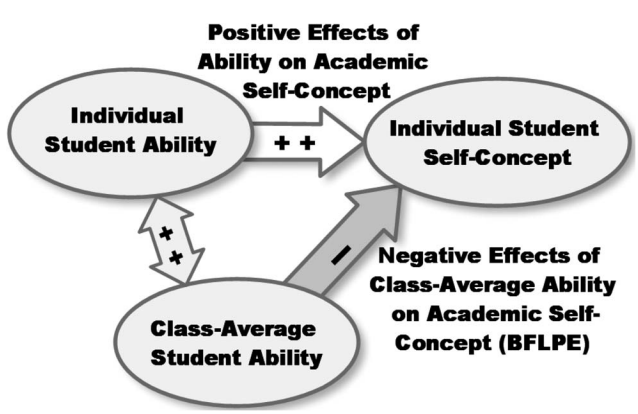

Figure 1. Conceptual model of the big-fish-little-pond effect (BFLPE).

in typical social psychology experiments involving introductory psychology students in contrived settings. Indeed, they argue that except for opting out altogether, it is difficult for students to avoid the relevance of achievement as a reference point within a class or school setting or the social comparisons provided by the academic accomplishments of their classmates (see also Marsh, 2007). Seaton, Marsh, et al. (2008) provided a theoretical rationale for how the BFLPE fits with the broader social comparison research literature, contrasting results for the imposed social comparisons and social comparison when students can freely choose their comparison targets. In support of the direct role of social comparison for the BFLPE, Huguet et al. (2009) demonstrated the BFLPE was largely eliminated after controlling pure measures of social comparison.

Extensive support for the BFLPE generalizes over student groups, subject domains, ASC instruments, and cultures (see reviews by Marsh, Seaton, et al., 2008; Seaton & Marsh, 2013). However, most BFLPE research has been based on high school students from Western developed countries (e.g., Australia, United States, Germany, Israel, France, the Netherlands, and the United Kingdom), as well as in Asian countries such as Hong Kong and Singapore.

Cross-Cultural Support for the BFLPE

In cross-cultural research there are two main orientations, one that focuses on tests of a priori hypotheses of cross-cultural differences and one that tests the replicability of existing theories in other cultures and seeks universal, panhuman theories (e.g., Parker et al., 2012; Segall, Lonner, & Berry, 1998, p. 1102). However, strong cross-cultural studies need to compare the results from at least two—and preferably many—countries based on comparable samples and the same measures; otherwise apparent cross-cultural differences are confounded with potential differences in the composition of samples and, perhaps, the appropriateness of materials. Addressing these challenges, there is strong support for the cross-cultural generalizability of the BFLPE for high school students, based on successive data collections of the Organisation for Economic Co-operation and Development Program for International Student Assessment (PISA) data: Marsh and Hau (2003) used the PISA 2000 data based on 103,558 fifteen-year-old students from 26 predominantly industrialized Western countries; Seaton, Marsh, and Craven (2009, 2010) used PISA 2003 (265,180 students, 10,221 schools, 41 countries), which included more collectivist and developing economies than PISA 2000; Nagengast and Marsh (2012) used the PISA 2006 database in the largest cross-cultural study of the BFLPE undertaken to date, and significantly extended the previous PISA studies. In summarizing the three BFLPE–PISA studies, Nagengast and Marsh reported that the effect of school-average achievement was negative in all but one of the 123 samples across the three studies, and significantly so in 114 samples. The average effect size across all 123 samples is −.223 (see detailed summary of this previous research in the supplemental materials, Table S3).

The overarching cross-cultural focus of our study was to evaluate the generalizability of the BFLPE in Middle Eastern Islamic countries (where there has been almost no research) to that found in Western countries (that is the basis of most research) and Asian countries (that has been the basis of some research). Although we note why these comparisons are potentially interesting, it was our

expectation that there would be reasonable support for the generalizability of the results. Indeed, Seaton et al. (2009) claimed support for the universality of the BFLPE as a panhuman theory based on PISA data. However, Schwartz and Bilsky (1990), as well as many others, observed, "Theories that aspire to universality . . . must be tested in numerous, culturally diverse samples" (p. 878). In this respect, one purpose of our study is to greatly expand the scope of tests of the BFLPE's generalizability beyond the set of PISA studies that have been the primary basis of cross-cultural tests of its universality.

PISA versus TIMSS. Although each of the successive PISA studies included a larger and more diverse sample of countries, there were important limitations that are the focus of the present investigation. For example, in their monograph on concerns related to PISA, Hopmann, Brinek, and Retzl (2007; see also Ertl, 2006) summarized a range of substantive, methodological, and policy-related concerns. These included the inappropriateness of the PISA model in respect of what is actually taught in many school systems, technical issues related to translation and scaling, problems with the sampling design, PISA's focus on literacy in testing mathematics, issues in relation to gender, and the league table ranking of countries based on PISA results. Potential concerns such as these dictate that cross-cultural BFLPE research based almost exclusively on PISA data should be cross-validated with data from different sources. Hence, it is surprising that there is apparently no cross-cultural BFLPE research based on the TIMSS data.

Systematic comparisons of results based on the TIMSS and PISA studies (e.g., American Institutes for Research, 2005; Hutchison & Schagen, 2007; National Center for Education Statistics, 2008; Neidorf, Binkley, Gattis, & Nohara, 2006; Wu, 2009) emphasize many similarities but also important differences between achievement tests in the two databases. In particular, PISA focuses more on the application of knowledge to "real life" problems, while TIMSS focuses on achievement more closely linked to school curriculums. Wu (2009) reported that these differences in item content explain in part why Western countries tended to perform better on PISA than TIMSS, while Eastern European and Asian countries tended to perform better on TIMSS than PISA.

A key distinction between TIMSS and PISA that is of particular relevance to the BFLPE lies in the differences in the way data have been collected. PISA samples schools, rather than classrooms, and then tests a random sample of 15-year-olds from each school, so that participants within the same school typically come from two-, three-, or four-year cohorts. Even at the school level, this sampling design complicates interpretation of frame-of-reference effects based on school-average achievement, which typically does not correspond to the achievement levels of students in any of the different year cohorts actually considered. In contrast, although TIMSS also samples schools, it measures all students from selected intact classrooms. However, although TIMSS is nationally representative of each country in relation to classes, there is typically only a single class selected from each school and this class may or may not be representative of the school as a whole. Hence, the appropriate unit of analysis with TIMSS is the classroom rather than the school.

Although the focus of both TIMSS and PISA has been on achievement scores, both databases include a range of psychosocial variables, including ASC responses that are the focus of the present investigation. For both PISA and TIMSS, considerable effort has been made to make the achievement scores comparable from one data collection to the next, although the rationale for testing achievement differs substantially in PISA and TIMSS. Whereas there has also been reasonable consistency in the items used to infer mathematics self-concept in the different TIMSS data collections (see discussion by Marsh et al., 2013; see also Method section for wording of TIMSS math self-concept items used here), this has not been the case for PISA. First of all, PISA typically only includes math self-concept responses when mathematics is the focus of the PISA data collection. Furthermore, the number and wording of items used in PISA to assess self-concept in the focal domain (math, science, or reading) varies substantially from one data collection to the next (for wording of PISA items in different data collections, see PISA website http://www.oecd.org/pisa/aboutpisa/). Finally, on the basis of these comparisons, it is also obvious that the items used to assess self-concept are clearly quite different across the TIMSS and PISA studies.

Local dominance effect. According to the local dominance effect (Zell & Alicke, 2009; see also Liem, Marsh, Martin, McInerney, & Yeung, 2013), the frame of reference—school versus classroom—is a potentially important consideration for BFLPE research. Zell and Alicke (2009; see also Alicke, Zell, & Bloom, 2010) provided support for the BFLPE by experimentally manipulating the frame of reference in relation to feedback given to participants about how their performances compared to others. When they pitted "local" against more "general" comparison standards, participants consistently used the most local comparison information available to them, even when they were told that the local comparison was not representative of the broader population and were provided with more appropriate normative comparison data. Hence, the class-average achievement based on the TIMSS data constitutes a more proximally relevant frame of reference than the school-average achievement based on PISA data, which is likely to be more locally dominant. In this respect, it is important to note that our study is apparently the first cross-cultural BFLPE study to be based on the classroom as the unit of analysis, rather than the school.

Developmental Support for the Generalizability of the BFLPE

For many developmental, educational, and psychological researchers, self-concepts are a "cornerstone of both social and emotional development" (Kagen, Moore, & Bredekamp, 1995, p. 18; see also Davis-Kean & Sandler, 2001; Marsh, Ellis, & Craven, 2002); self-concepts develop early in childhood and, once established, they are enduring (e.g., Eder & Mangelsdorf, 1997). The development of self-concept is therefore emphasized in many early childhood programs (e.g., Fantuzzo, McDermott, Manz, Hampton, & Burdick, 1996).

Hattie (1992; Hattie & Marsh, 1996; see also Eccles, Wigfield, Harold, & Blumenfeld, 1993; Harter, 1999, 2006, 2012; Marsh, Craven, & Debus, 1998) reviewed theoretical and empirical support for stages of growth in the development of self-concept, arguing against the notion of fixed stages that all persons must pass through. Instead, he posited seven parallel developments that are relevant to self-concept formation: (a) children distinguish self and others, (b) children distinguish self and the environment, (c) changes in major reference groups lead to changes in expectations,

(d) attributions are made to salient personal and social or external sources, (e) cognitive processing capacities develop, (f) children develop particular cultural values, and (g) children develop strategies for confirmation and disconfirmation of self-referent information. Thus, with age and development, young children increasingly integrate information from their immediate environment into their self-concept formation. This is particularly relevant to the present investigation, emphasizing the integration of external frames of reference and social comparison into self-concept formation.

Indeed, many authors (Chapman & Tunmer, 1995; Eccles et al., 1993; Harter, 1999; Marsh, 1989; Marsh & Craven, 1997; Skaalvik & Hagtvet, 1990; Wigfield & Eccles, 1992; Wigfield et al., 1997) have offered a developmental perspective on the relation between ASC and academic achievement. For example, Marsh (1989, 1990; Marsh et al., 1998) proposed that the self-concepts of very young children are very positive and are not highly correlated with external indicators (e.g., skills, accomplishments, achievement, self-concepts inferred by significant others) but that with increasing life experience, children learn their relative strengths and weaknesses, so that specific self-concept domains become more differentiated and more highly correlated with external indicators. Marsh et al. (1998) showed that reliability, stability, and factor structure of self-concept scales improve with age (children 5–8 years of age). In addition, consistent with the proposal that children's self-perceptions become more realistic with age, self-ratings of older children were more correlated with inferred self-concept ratings by their teachers.

In summary, there is good developmental theory for the prediction that with age and development ASC becomes more closely related to external criteria, including academic achievement and perceptions of others. From this, it is reasonable to speculate that the BFLPE would also become stronger with age and development, but there is little or no empirical evidence against which to evaluate this supposition. Testing this generalizability of the BFLPE over primary and secondary students is the major focus of our study.

TIMSS 2007: Background to the Present Investigation

An important aspect of TIMSS is collection of data from two age cohorts (corresponding to fourth and eighth grades), providing a unique developmental perspective on cross-cultural studies of the BFLPE. Included in the present investigation are fourth- and eighth-grade classes (see Table 1 for more detail) in six Western countries (Australia, England, Italy, Norway, Scotland, and United States), four Asian countries (Hong Kong, Japan, Singapore, and Taiwan), and three Middle Eastern Islamic countries (Iran, Kuwait, and Tunisia) where both mathematics and science were taught as an integrated subject, and where data were available for both fourth- and eighth-grade cohorts.

In line with the substantial body of BFLPE, we predict that there will be good support for the developmental generalizability of the BFLPE across the two matched age cohorts, and for the cross-cultural generalizability of the BFLPE across the 13 countries. In keeping with developmental models of self-concept (and limited empirical support), we posit that relations between ASC and achievement will be significant for all 26 (2 age cohorts × 13 countries) groups but will be stronger for the older age cohort. Of central importance, we predict that the negative effect of class-average achievement on ASC—the BFLPE—will also be significant across all 26 groups. However, we further surmise that the BFLPE will be stronger for the older cohort, in that developmental models posit that social comparison and normative processes in the formation of self-concept grow stronger over this developmental period; but we recognize that there is limited empirical support for this prediction. We leave as a research question whether there are substantively meaningful interactions between country and age cohort differences; alternatively, the extent to which age cohort effects generalize across countries.

Table 1
Reliability of the Trends in International Mathematics and Science Study Math and Science Motivation Constructs Used in This Study

Country	Cohort	Number of Students	Number of Classes	Number of Schools	Reliability
Western countries					
Australia	4	4,108	316	228	.747
	8	4,119	238	227	.809
England	4	4,316	233	142	.753
	8	4,046	238	136	.795
Italy	4	4,470	323	169	.687
	8	4,408	287	169	.841
Norway	4	4,108	290	144	.677
	8	4,748	264	138	.805
Scotland	4	3,929	252	138	.723
	8	4,213	244	128	.770
United States	4	7,896	515	256	.763
	8	7,636	510	238	.838
Total	4	28,827	1,929	1,077	.725
	8	29,170	1,781	1,036	.810
Asian countries					
Taiwan	4	4,131	174	149	.735
	8	4,046	153	149	.838
Hong Kong	4	3,791	147	125	.717
	8	3,527	120	119	.803
Japan	4	4,487	189	147	.762
	8	5,625	169	145	.777
Singapore	4	5,041	354	176	.757
	8	4,754	326	163	.825
Total	4	17,450	864	597	.743
	8	17,952	768	576	.811
Middle Eastern Islamic countries					
Iran	4	3,833	224	223	.734
	8	3,981	208	207	.744
Kuwait	4	3,803	181	149	.351
	8	4,091	158	157	.589
Tunisia	4	4,134	217	149	.450
	8	4,080	169	149	.729
Total	4	11,770	622	521	.512
	8	12,152	535	513	.687
All countries					
Grade	4	58,047	3,415	2,195	.681
	8	59,274	3,084	2,125	.781
Total		117,321	6,499	4,320	.731

Note. Reliabilities are expressed as Cronbach's alpha estimates.

Method

Participants

TIMSS 2007 (Olson, Martin, & Mullis, 2008) assessed the competencies in mathematics for nationally representative samples of students from participating countries (for more details about the processes underlying the development of the TIMSS 2007 instruments, translation of materials, sampling, data collection, scaling, and data analysis, see the TIMSS 2007 technical report by Olson et al., 2008). The basic design is a two-stage cluster design that consists of sampling schools and intact classrooms from the target grade in the school. Included in the present investigation are 117,321 students from 6,499 fourth- and eighth-grade classes representing 13 countries (see Table 1 for more detail). In all countries, the materials were administered near the end of the school year (typically October or November in the Southern Hemisphere and April to June in the Northern Hemisphere). For purposes of convenience and consistency with TIMSS 2008, we refer to the fourth-grade cohort (typically 9–11 years of age with 4 years of formal schooling) as primary school children and the eighth-grade cohort (typically 13–15 years of age with 8 years of formal schooling) as secondary school adolescents, but realize that this terminology is not completely consistent across all countries and school systems.

TIMSS (Olson et al., 2008) used item response theory to scale student achievement scores based on a mixture of constructed response and multiple-choice items representing algebra, data and chance, number, and geometry for eighth-grade students and number, geometric shapes and measures, and data display for fourth-grade students. Students in both age cohorts responded to items designed to measure math self-concept (MSC) on the same classic Likert (agree–disagree) response scale; two of the self-concept items had the same wording in the two age cohorts, but there were minor wording changes for the other two self-concept items (see supplemental materials): "I usually do well in math"; "Math is harder for me than for many of my classmates"; "I am just not good at science"; "I learn things quickly in math/science."

Data Analysis

All analyses, conducted with Mplus 7.0 (Muthén & Muthén, 2013), consisted of multilevel confirmatory factory analyses (CFAs) and structural equation models (SEMs) based on the Mplus robust maximum likelihood estimator, with standard errors and tests of fit that were robust in relation to nonnormality of observations and the use of Likert responses where there were at least four or more response categories, particularly where nonnormality was not excessive (e.g., Beauducel & Herzberg, 2006; DiStefano, 2002; Dolan, 1994; Muthén & Kaplan, 1985). Maximum likelihood estimation is also robust to the nonindependence of the observations when used in conjunction with a design-based correction (Mplus's complex design option; Muthén & Muthén, 2013). All analyses were based on TIMSS's HOUWGT weighting variable that incorporates three components related to sampling of the school, class, and student, respectively, and three associated with nonparticipation at the level of the school, class, and student. For present purposes, the 26 (13 countries × 2 age cohorts) groups were treated as grouping variables that were the basis of the multigroup analyses, whereas the class and school identification variable was treated as a clustering variable to control for the clustered sample (using the complex design option and robust maximum likelihood options in Mplus; Muthén & Muthén, 2013).

In the TIMSS 2007 database, the achievement tests for each student are reported as five plausible values (Olson et al., 2008)—numbers drawn randomly from the distribution of scores that could be reasonably assigned to each student. Although the amount of missing data was small (an average of less than 2% per item), we used full-information maximum likelihood estimation to control for missing data, noting that this had been done separately for each of the five data sets based on different plausible values and then combined using the Rubin (1987; Schafer, 1997) strategy to obtain unbiased parameter estimates, standard errors, and goodness-of fit statistics.

Comparison of results across different countries, and age cohorts, requires strong assumptions about the invariance of the factor structure. If the underlying factors differ fundamentally in different groups, then there is no basis for interpreting observed differences (the "apples and oranges" problem; see Millsap, 2011). Here we initially consider invariance across the 26 (13 countries × 2 age cohorts) groups. In applied SEM research—particularly for large sample sizes as in TIMSS—there is a predominant focus on indices that are sample size independent (e.g., Hu & Bentler, 1999; Marsh, Balla, & McDonald, 1988; Marsh, Hau, & Grayson, 2005; Marsh, Hau, & Wen, 2004). The Tucker–Lewis index (TLI) and the comparative fit index (CFI) vary along a 0–1 continuum, and values greater than .90 and .95 typically reflect acceptable and excellent fit to the data, respectively. For the root-mean-square error of approximation (RMSEA), values of less than .05 and .08 reflect a close fit and a minimally acceptable fit to the data, respectively. However, for purposes of model comparison, comparison of the relative fit of models imposing more or fewer invariance constraints, Cheung and Rensvold (2002) and Chen (2007) suggest that if the decrease in fit for the more parsimonious model is less than .01 for incremental fit indices like the CFI, then there is reasonable support for the more parsimonious model. Chen suggests that when the RMSEA increases by less than .015, there is support for the more constrained model. However, these guidelines are based on simulated data studies and practice typically involving only two or a small number of groups, and might not be fully applicable to studies like the present investigation, based on 26 groups. Hence, it is important to emphasize that these are only rough guidelines (Marsh, Hau, & Wen, 2004), and it is recommended that applied researchers use an eclectic approach based on subjective integration of a variety of different indices—including the chi-square, detailed evaluations of parameter estimates in relation to theory, a priori predictions, common sense, and alternative models specifically designed to evaluate goodness of fit in relation to key issues. This is consistent with the approach we used here.

Latent contextual effect models: Substantive-methodological synergy. Only recently has BFLPE research integrated the application of multilevel models (e.g., students nested within classes) with the use of latent variable models (with multiple indicators of latent constructs, the multiple MSC items) and multiple group analyses to compare results across countries (see Lüdtke, Marsh, Robitzsch, & Trautwein; Lüdtke et al., 2008; Marsh, Lüdtke, et al., 2009; Nagengast & Marsh, 2012). In the present application of this evolving statistical approach (see Figure 2), we used manifest aggregation to form the class-average measure of achievement such that

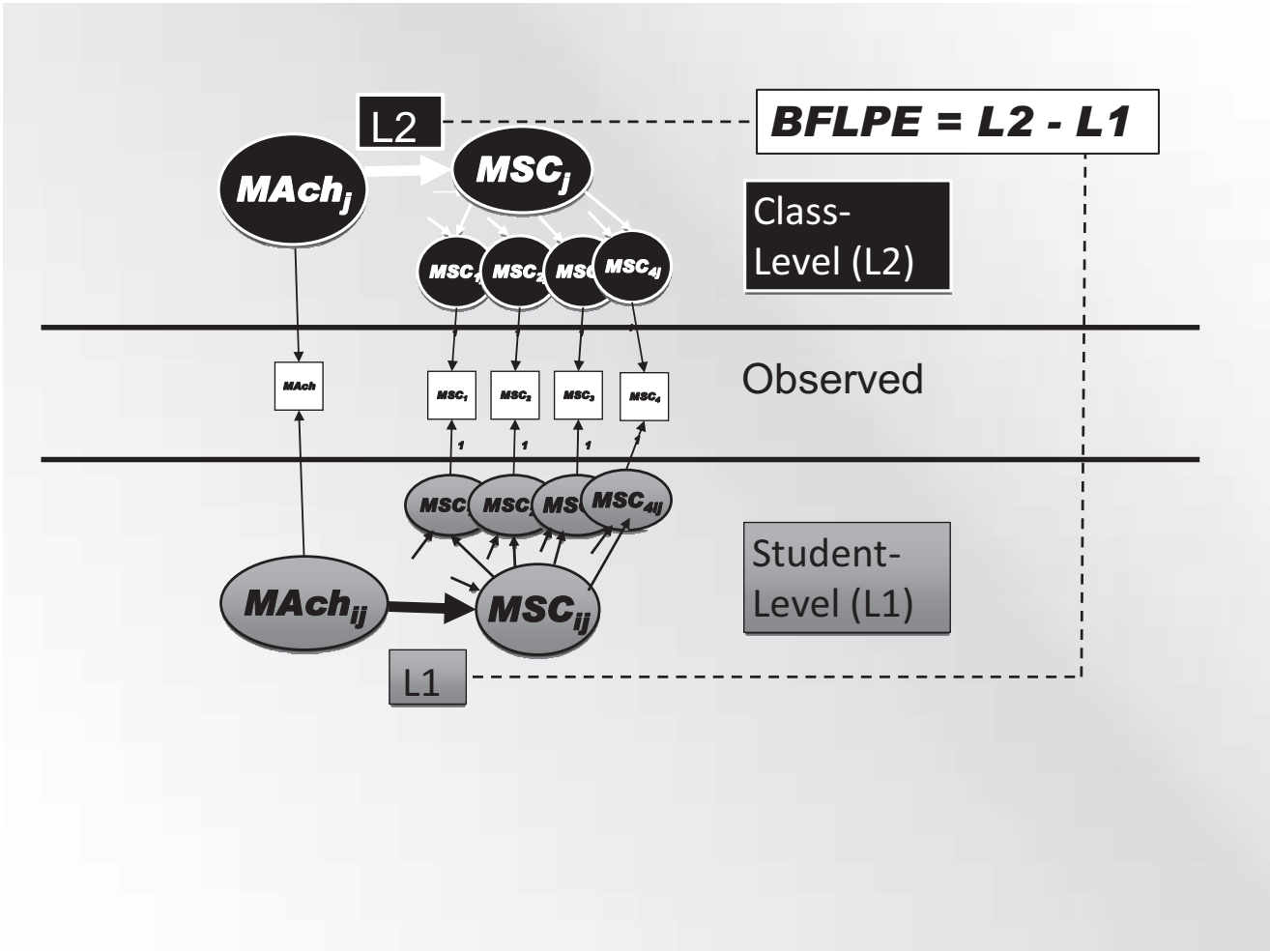

Figure 2. Multilevel depiction of the big-fish-little-pond effect (BFLPE). MAch = math achievement; MSC = math self-concept.

$$\text{Self-concept} = \beta_0 + \beta_1\,(\text{achievement}) + r$$
$$\beta_0 = \gamma_{00} + \gamma_{01}\,(\text{mean}_{\text{achievement}}) + \mu_0, \quad (1)$$

where β_0 is a random intercept and β_1 is the effect of individual student achievement on self-concept; γ_{01} represents the variation in β_0 that is explained by school-average achievement; r and μ_0 are residual terms. For TIMSS data, intact classes were sampled so that the sampling ratio approached 1.0 and so sampling error was minimal. Indeed, the use of latent aggregation (and the use of within-class achievement variation to estimate sampling error) would overcorrect BFLPE estimates (see Marsh, Lüdtke, et al., 2009; although the size of the bias would be small because of the substantial sample sizes).

The contextual effects models were estimated with the reflective aggregation procedure in Mplus (Muthén & Muthén, 2013) that uses implicit group-mean centering of all L1 variables. This implies that the partial regression weights associated with L1 variables reflect L1 effects, while the partial regression weights associated with L2 variables reflect L2 effects that are not controlled for L1 differences (Enders & Tofighi, 2007; Kreft, de Leeuw, & Aiken, 1995). Estimates of contextual effects, that represent the effect of L2 variables after controlling for L1-differences, can be obtained by subtracting the L1 effect from the L2 effect (Enders & Tofighi, 2007; Kreft et al., 1995):

$$\beta_{\text{context}} = \beta_{L2} - \beta_{L1}, \quad (2)$$

where β_{L2} is the L2 effect, β_{L1} is the L1 effect, and β_{context} is the contextual effect (see Figure 2). The standard error for the contextual effect was obtained with the multivariate delta method (see Raykov & Marcoulides, 2004).

In order to facilitate comparisons with previous research, effect sizes (ESs) for the BFLPE (the effect of class-average achievement on MSC after controlling for individual student achievement) were calculated according to the recommendations of Marsh, Lüdtke, et al. (2009; Nagengast & Marsh, 2012) by the following formula:

$$\text{BFLPE ES} = 2 * \beta * \sigma_{\text{pred}}/\sigma_y, \quad (3)$$

where β is the unstandardized regression coefficient, σ_{pred} is the standard deviation of the predictor variable (achievement), and σ_y

is the standard deviation of the outcome variable (self-concept), resulting in an ES metric that is common across countries. This ES is comparable to Cohen's *d* (Cohen, 1988), reflecting differences based on classes 1 standard deviation above the mean and 1 standard deviation below the mean. For MSC, students in all 26 (13 countries × 2 cohorts) groups completed the same items, and so it was appropriate to standardize MSC responses in relation to a standard deviation that was common across all 26 groups. However, for math achievement (MAch), the tests were completely different for the two age cohorts, so that we standardized the achievement scores to have $M = 0$ and $SD = 1$ across the 13 countries within each of the two age cohorts.

In the decomposition of group (13 countries × 2 age cohorts) into variance components and more detailed factorial (analysis-of-variance-like) contrasts, we relied heavily on the flexibility of the "model constraint" function in Mplus. The resulting tests of statistical significance based on these model constraints were based on the delta method (Muthén & Muthén, 2013). Thus for example, we used these constraints to obtain analysis-of-variance-like estimates of the statistical significance and proportion of variation in the relations of MSC with student level (L1) achievement and class-average (L2) achievement that was explained by the 13 countries (and three groups of countries: Western, Asian, Middle Eastern Islamic), two age cohorts (Grade 4 versus Grade 8), and Age Cohort × Country interactions. These were followed by more specific tests of a priori hypotheses. This evolving methodology—combining the flexibility typically associated with analyses of manifest variables with latent variable models—is apparently a new contribution.

Preliminary Results

In preliminary analyses we estimated the average reliability of the MSC score and evaluated the a priori factor structure for these responses based on Marsh et al. (2013; see supplemental materials for more detail). Due in part to the brevity of the four-item MSC scale, reliability estimates (see Table 1) sometimes reached a desirable standard of .80, but in other cases fell below acceptable values of .70 or even .60. Reliability estimates were systematically higher for the older age cohort (mean $\alpha = .781$) than the younger cohort (mean $\alpha = .681$), and substantially lower in the Middle Eastern Islamic countries than in the Western or Asian countries. These systematic differences in reliability make problematic comparisons based on manifest scale or composite scores, and support the need to consider latent variable models that control for unreliability.

Our a priori factor model (following from Marsh et al., 2013; see supplemental materials) is a simple model in which the four self-concept items are associated with one latent self-concept factor, MAch is a single-item variable (represented by TIMSS's five sets of plausible values that control for unreliability), and there is a negative-item method effect represented by a correlated uniqueness between the two negatively worded self-concept items. This model was supported based on a series of single-level multigroup (using the Mplus complex design to control for clustering of students within classes and schools) tests of invariance over 26 (2 cohorts × 13 countries) groups. Next we tested multilevel–multigroup CFA models demonstrating the invariance of factor loadings within and across student (L1) and class (L2) levels as well as the 26 groups. Subsequent results supported the highly constrained model, in which all factor loadings were constrained to be the same across all 26 groups at both the student and class levels (CFI = .956, TLI = .941, RMSEA = .054; see supplemental materials for the Mplus syntax used to test this model), that was the basis of subsequent results.

Results

Support for the BFLPE requires that the effect of individual (L1) achievement on MSC is positive, while the effect of class-average (L2) achievement is negative (see Figure 1). The standardized path coefficients between individual student achievement and MSC are significantly positive in all 26 (13 countries × 2 cohorts) groups ($M = .592$, $SE = .005$; see Table 2). In contrast, the BFLPEs (ESs for the negative effect of class-average achievement on MSC) are significantly negative in all 26 groups ($M = -.377$, $SE = .012$; see Table 2). These results provide strong support for the generalizability across countries and across age cohorts. We now address substantively important developmental and cross-cultural issues, evaluating how these effects vary as a function of age cohort, country, and their interaction.

Relations Between Student Level (L1) Achievement and Self-Concept

Averaged across all countries and age cohorts, achievement and MSC are positively correlated ($M = .592$, $SE = .005$; see Table 2). Next, we decomposed variance estimates into contrast tests of differences associated with the 13 countries, the two age cohorts, and their interactions; and estimated variance components for each of these differences (sums of squares and variance components in Table 2). Estimates for all 26 groups, the mean estimate for each country, the mean estimate for each cohort, and the means for each of the three country groupings are all significant and positive. The variance components associated with each effect—along with an inspection of the values for each of the 26 (2 cohorts × 13 countries) groups—provide an indication of the sizes of the effects and how well they generalize over age cohorts and countries.

Cohort effects. The size of the relation between MAch and MSC, averaged across all countries, is substantially larger for the older cohort ($M = .692$, $SE = .008$) than for the younger cohort ($M = .492$, $SE = .006$). However, interpretations of cohort differences are complicated by Cohort × Country interactions, suggesting that cohort differences vary for different countries. The positive relations are substantially larger for secondary than primary students (i.e., difference scores in Table 2 are significantly positive) in all Western and in all Islamic countries, but not in all the Asian countries. Although the mean estimate across the Asian countries is significantly ($p < .05$) higher for the older cohort, the difference is small and inconsistent across the Asian countries. Only in Singapore is the positive relation larger for secondary than primary students, and in Taiwan the positive relation is significantly larger for primary than for secondary students (cohort differences are not significant in Hong Kong and Japan).

Country differences. Although the relation between self-concept and achievement is significantly positive for both cohorts in each of the 13 countries (see Table 2), estimates are more positive in the Western ($M = .627$, $SE = .008$) and Asian ($M =$

Table 2
Effects of Individual Student Achievement and Class-Average Achievement on Math Self-Concept Broken Down by 13 Countries × 2 Age Cohorts

Country	Age cohort	Individual achievement	Class-average achievement (BFLPE effect size)
Western countries			
Australia	4	.583 (.021)	−.358 (.042)
	8	.914 (.032)	−.627 (.063)
	Diff	.331 (.038)	−.269 (.075)
	Total	.749 (.019)	−.493 (.038)
England	4	.460 (.018)	−.294 (.048)
	8	.629 (.035)	−.359 (.051)
	Diff	.169 (.039)	−.065 (.069)*
	Total	.544 (.019)	−.327 (.035)
Italy	4	.481 (.019)	−.482 (.041)
	8	.832 (.020)	−.907 (.068)
	Diff	.351 (.028)	−.425 (.079)
	Total	.656 (.014)	−.694 (.039)
Norway	4	.369 (.019)	−.134 (.054)
	8	.937 (.022)	−.527 (.086)
	Diff	.568 (.030)	−.393 (.104)
	Total	.653 (.014)	−.331 (.050)
Scotland	4	.364 (.019)	−.418 (.072)
	8	.563 (.032)	−.282 (.058)
	Diff	.199 (.036)	.137 (.093)*
	Total	.463 (.019)	−.350 (.046)
United States	4	.631 (.018)	−.352 (.038)
	8	.766 (.025)	−.502 (.050)
	Diff	.135 (.033)	−.150 (.065)
	Total	.699 (.014)	−.427 (.030)
Mean Western	4	.481 (.009)	−.340 (.021)
	8	.774 (.013)	−.534 (.026)
	Diff	.292 (.016)	−.194 (.034)
	Total	.627 (.008)	−.437 (.016)
Asian countries			
Hong Kong	4	.609 (.032)	−.441 (.062)
	8	.636 (.034)	−.549 (.051)
	Diff	−.027 (.054)*	−.107 (.085)*
	Total	.623 (.020)	−.495 (.038)
Japan	4	.578 (.019)	−.247 (.078)
	8	.574 (.019)	−.482 (.068)
	Diff	.004 (.027)*	−.236 (.106)
	Total	.576 (.013)	−.364 (.051)
Taiwan	4	.688 (.026)	−.475 (.077)
	8	.581 (.019)	−.180 (.056)
	Diff	−.107 (.028)	.295 (.095)
	Total	.635 (.018)	−.327 (.047)
Singapore	4	.589 (.019)	−.211 (.040)
	8	.828 (.033)	−.585 (.062)
	Diff	.239 (.036)	−.374 (.074)
	Total	.708 (.020)	−.398 (.037)
Mean Asia	4	.616 (.011)	−.343 (.033)
	8	.655 (.016)	−.449 (.031)
	Diff	.039 (.018)	−.106 (.047)
	Total	.635 (.010)	−.396 (.022)
Middle Eastern Islamic countries			
Iran	4	.510 (.022)	−.175 (.044)
	8	.580 (.023)	−.362 (.047)
	Diff	.069 (.031)	−.186 (.064)
	Total	.545 (.016)	−.268 (.032)
Kuwait	4	.236 (.017)	−.089 (.038)
	8	.457 (.023)	−.342 (.065)
	Diff	.221 (.027)	−.252 (.076)
	Total	.347 (.015)	−.216 (.037)
Tunisia	4	.300 (.014)	−.117 (.048)
	8	.703 (.026)	−.314 (.102)
	Diff	.403 (.030)	−.197 (.112)*
	Total	.502 (.015)	−.216 (.057)
Mean Islam	4	.349 (.010)	−.127 (.025)
	8	.580 (.014)	−.339 (.043)
	Diff	.231 (.016)	−.212 (.050)
	Total	.464 (.009)	−.233 (.025)
Mean across all 26 (13 countries × 2 cohorts) groups			
Total	4	.492 (.006)	−.292 (.015)
	8	.692 (.008)	−.463 (.018)
	Diff	.200 (.010)	−.171 (.023)
	Total	.592 (.005)	−.377 (.012)
Sums of squares (SS) and variance components (VC)[a]			
SS cohort		.260 (.026)	.210 (.053)
VC		.067	.013
SS country		.300 (.033)	.419 (.077)
VC		.077	.026
SS interaction		.205 (.021)	.226 (.060)
VC		.053	.014

Note. For each country, the three country groupings, and the total across all 26 (13 countries × 2 cohorts) groups, results are presented for each age cohort (fourth grade and eighth grade), the difference (diff) between age cohorts, and the total across age cohorts. For each estimate there is a standard error (in parentheses) that can be used to assess statistical significance (i.e., estimates divided by their standard error that are greater than 1.96 are statistically significant at $p < .05$). For individual student achievement, the estimates are the standardized path coefficients, while for the big-fish-little-pond effect (BFLPE) the estimates are effect sizes.
[a] Effects across the 26 groups were decomposed to assess the main effects of differences due to the 13 countries, the two age cohorts, and their interaction (sums of squares and variance components).
* Estimates that are not statistically significant in the predicted direction; all other estimates are statistically significant at $p < .05$.

.635, SE = .010) than in the Islamic (M = .464, SE = .009) countries. However, results for individual countries are not entirely consistent, even within each of the three country classifications, and these differences interact significantly with cohort. For the younger cohort, the relations are substantially more positive in Asian countries (M = .616, SE = .011) than in the Western (M = .481, SE = .009) and particularly the Islamic (M = .349, SE = .010) countries. However, for the older cohort, relations are substantially more positive in the Western countries (M = .774, SE = .013) than in Asian (M = .616, SE = .011) and particularly in Islamic (M = .580, SE = .014) countries. Averaged across cohorts, the estimates are most positive in Australia, Singapore, and the United States, but are also higher in Italy, Norway, Taiwan, Hong Kong, and Japan than in any of the three Islamic countries.

The BFLPE: The Negative Effects of Class-Average Ability on MSC

Averaged across all countries and age cohorts, class-average achievement has a negative effect on MSC (mean BFLPE ES = −.377, SE = .012; see Table 2). Although the BFLPE is significantly negative for each of the 26 (2 cohorts × 13 countries)

groups, its size does vary significantly with cohort, country, and their interaction, as demonstrated by variance components—along with an inspection of the values for each of the 26 groups.

Cohort effects. Because there have been no previous cross-cultural studies of the BFLPE with primary school students, the most important result of our study is that in all 13 countries the BFLPE is statistically significant and negative for the younger cohort, as well as the older cohort (see Table 2). An important contribution of our study is the finding—consistent with predictions—that the BFLPE is significantly larger in the eighth-grade cohort (mean BFLPE ES = −.463, *SE* = .018) than in the fourth-grade cohort (mean BFLPE ES = −.283, *SE* = .015). Furthermore, the mean BFLPE ESs are significantly more negative for the older cohort in each of the three country groups, and these cohort differences do not differ significantly from each other (West: −.194, *SE* = .065; Asian: −.106, *SE* = .047; Islamic: −.212, *SE* = .050). Again, however, interpretations of cohort differences are complicated by Cohort × Country interactions. For four countries (England, Scotland, Hong Kong, and Tunisia) the BFLPE did not differ significantly as a function of cohort, while in one country (Taiwan) the BFLPE was significantly more negative for the younger cohort.

Country differences. The BFLPE is significantly negative for both cohorts in all 13 countries, but there are differences between countries (see Table 2). Across the cohorts, the BFLPE is more negative in the Western (mean BFLPE ES = −.437, *SE* = .016) and Asian (mean BFLPE ES = −.396, *SE* = .022) countries than in Islamic (mean BFLPE ES = −.233, *SE* = .025) countries. As already noted, the BFLPE is noticeably smaller in the younger cohort of Islamic countries (−.127, *SE* = .025). The BFLPE is particularly large in Italy (mean BFLPE ES = −.694, *SE* = .039), but is also very substantial in Hong Kong (mean BFLPE ES = −.495, *SE* = .038), Australia (mean BFLPE ES = −.493, *SE* = .038), and, to a lesser extent, the United States (mean BFLPE ES = −.427, *SE* = .030) and Singapore (mean BFLPE ES = −.398, *SE* = .037). The BFLPE was smallest in Tunisia (mean BFLPE ES = −.216, *SE* = .057) and Kuwait (mean BFLPE ES = −.216, *SE* = .037), particularly for the younger cohort.

Discussion

Substantively and theoretically, the most important result of the present investigation is that the BFLPE—the negative effect of class-average achievement on MSC—is statistically significant and generalizes across both age cohorts in all 13 countries, providing good support for its developmental and cross-national generalizability. This is important because this is the only large-scale cross-cultural study to compare the BFLPE across matched samples of primary and secondary students from a broad array of different countries.

Our study is also the first large-scale cross-cultural study of the BFLPE not based on PISA. Importantly, the consistency of the BFLPEs for both cohorts for the TIMSS data in our study is even stronger than in previous cross-cultural studies based on PISA data. Thus, the average BFLPE ES across 123 samples based on PISA data (59 countries sampled in one or more data collections in PISA 2000, PISA 2003, and PISA 2006) is −.223, while the average BFLPE ES across 24 samples (12 countries × 2 age cohorts) in the present study is −.377. Furthermore this general trend is reasonably consistent across overlapping countries that participated in both PISA and TIMSS. This might seem surprising, in that PISA data is based on somewhat older students—15-year-olds—than even the oldest TIMSS cohort, and our results, consistent with a priori development predictions, show that the BFLPE is more negative for older students (−.292 for Year 4, −.426 for Year 8). However, these findings are consistent with our a priori predictions based on the local dominance effect when comparing results based on school-average achievement (PISA) and class-average achievement (TIMSS). Nevertheless, there are a number of critical differences between TIMSS and PISA sampling designs that might explain, in part, these differences but also dictate caution in interpretation of the results:

- The nature of the standardized achievement tests is more closely related to the academic curriculum in TIMSS than in PISA, so that the frame of reference based on class-average TIMSS is more closely associated with the achievement results (e.g., school grades, teacher feedback) that are actually experienced by these students.
- The sampling unit for PISA is the whole school, while that of TIMSS is the individual class. Although both are relevant, there is some theoretical and empirical research (see earlier discussion of the local dominance effect; e.g., Liem et al. 2013; Zell & Alicke, 2009) suggesting that the more proximal frame of reference associated with the class is stronger—more locally dominant—than that associated with the whole school, particularly if there is streaming or tracking within schools, so that there are systematic differences in achievement levels for different tracks within schools.
- TIMSS samples intact classes that almost always represent a single-year group, whereas PISA samples 15-year-olds and typically includes two, three, or even four year groups within a given school. Thus, BFLPE interpretations for PISA are complicated in that the school-average ability estimate is an aggregation of test scores across multiple-year groups that do not correspond to any one of the year groups actually considered. Also, school-average ability estimates in PISA studies are typically based on a relatively small proportion of the 15-year-olds in the school, so that at least moderate amounts of sampling error in the school-average estimates are likely; the school-average estimate is a sample mean estimate of the true school-average value if all 15-year-olds in the school are tested. Although recent advances in contextual models used to assess the BFLPE are able to control for sampling error (Marsh et al., 2009; Nagengast & Marsh, 2012), this is typically at the expense of statistical power. In contrast, class-average estimates of achievement based on TIMSS scores are based on intact classes, so that there is little or no sampling error.

In summary, our results demonstrate that the size of the BFLPE is systematically larger for high school students than primary school students, although the BFLPE is clearly evident in both age cohorts. Our findings also suggest, consistent with the local dominance effect, that the BFLPE is substantially more negative for class-average achievement, as in TIMSS data, than for school-average achievement, as in PISA data. However, neither PISA nor TIMSS is ideal for testing this difference in that neither of these international comparisons allows researchers to distinguish properly between the effects of class- and school-average achievement in a more appropriate three-level (L1 = students, L2 = classes, L3 = schools) model. Furthermore, although it is likely that future

research will be able to address this issue with data from individual countries, it seems unlikely that that future research will be able to test the cross-cultural generalizability of these results with data as comprehensive as the PISA and TIMSS data. However, a comprehensive meta-analysis of BFLPE studies (that included PISA and TIMSS studies as well as the large number of studies done in individual countries) would be a useful addition to the literature.

Our study is also apparently the first to specifically compare BFLPE results in a sample of Middle Eastern Islamic countries with those from Asian and Western countries, which have been the basis of most BFLPE research. Indeed, only one of these Islamic countries in our study (Tunisia) had participated in PISA. Although the BFLPE was statistically significant for both age cohorts in all the Islamic countries, it was significantly smaller—particularly in the younger age cohort. In line with earlier research in Arab and Islamic countries (Marsh et al., 2013; see also Abu-Hilal & Bahri, 2000), we also found that relations between L1 achievement and MSC were significantly smaller in the Islamic countries—again, particularly for the younger age cohort. These authors previously speculated that students from these countries do not receive as much evaluative feedback about their achievement as Western and Asian students, and are not socialized in such a way as to critically evaluate their academic skills in relation to classmates. Indeed, consistent with speculations by Abu-Hilal and Bahri (2000) that self-concept formation of ASC and its relation with achievement in Middle Eastern Islamic middle school students was similar to that found in younger students from Western countries, support for the BFLPE for eighth-grade Middle Eastern Islamic students is similar to that found for the fourth-grade cohort in the Western and Asian countries (for further discussion, see Abu-Hilal, 2001; Abu-Hilal & Aal-Hussain, 1997; Abu-Hilal & Bahri, 2000).

Strengths, Limitations, and Directions for Further Research

Important strengths of our study include the use of large, nationally representative samples of primary and secondary school students from culturally diverse countries who were tested with standardized materials under standardized conditions; the integration of CFA, SEM, and multiple-group and multilevel modeling into a single analytic framework; and decomposition of differences in the BFLPEs associated with age cohort, country, and their interaction. In these respects, our study is a strong exemplar of the methodological-substantive synergies that apply evolving statistical methodology to substantively and theoretically important issues that have policy and practice implications. Nevertheless, as is always the case, there are important limitations that may provide the basis of further research.

Reliance on cross-sectional data for only two age cohorts. Reliance on only two cross-sectional age cohorts requires additional caveats in the interpretation of the results from a developmental perspective. For example, the apparent differences as a function of age might also be a function of birth cohort effects, and we were not able to evaluate how consistent the effects of age were for different individuals. Nevertheless, there are also some limitations with longitudinal data (e.g., generalizability of the results to other age cohorts; complications in sampling designs, missing data, representativeness of data within each country, and comparability across countries—particularly in relation to tracking students from primary to secondary schools). Ultimately, the "best" developmental description of how the BFLPE must incorporate findings from both cross-sectional and longitudinal studies, more fully evaluate developmental aspects of the BFLPE, and use a wider range of ages based on a combination of multicohort and longitudinal data.

Assumptions of causality and underlying processes. Support for the BFLPE—and contextual models more generally—is largely based on cross-sectional, correlational studies, so that causal interpretations should be offered tentatively and interpreted cautiously. In particular, the "third variable" problem is always a threat to contextual studies that do not involve random assignment, but Marsh, Hau, and Craven (2004; Marsh, Seaton, et al., 2008) argue that this is an unlikely counterexplanation of BFLPE results, in that most potential third variables (resources, per student expenditures, socioeconomic status, teacher qualifications, etc.) are positively related to class- or school-average achievement, so that controlling for them would typically increase the size of the BFLPE. Fortunately, there is now a growing body of BFLPE research using various combinations of longitudinal, quasi-experimental, and true experimental designs that all support the BFLPE (see Marsh, 2007; Marsh, Seaton, et al., 2008; Nagengast & Marsh, 2012). Quasi-experimental, longitudinal studies (e.g., Marsh, Kong, & Hau, 2000) show that students' ASC declines when students shift from mixed-ability schools to academically selective schools over time (based on pre- and posttest comparisons) and compared to students matched on academic ability who continue to attend mixed-ability schools. There is support for the BFLPE in studies where achievement is based on tests administered before students began high school (e.g., Marsh et al., 2000). Extended longitudinal studies (Marsh et al., 2000; Marsh, Trautwein, Lüdtke, Baumert, & Köller, 2007) show that the BFLPE grows more negative the longer students attend a selective school and is maintained even 2 and 4 years after graduation from high school. Also, there is good support for the theoretical underpinnings of the BFLPE, as it is largely limited to academic components of self-concept and nearly unrelated to nonacademic components of self-concept and self-esteem (Marsh, 1987; Marsh & Parker, 1984). However, further longitudinal and intervention studies would be useful to bolster the case for mediation of the effects of L1 and L2 achievement on subsequent achievement and educational attainment by ASC.

Also implicit in the BFLPE is the assumption that the direction of causal ordering is from class- or school-average ability to ASC. Although apparently reasonable, this implicit causal ordering cannot easily be tested with cross-sectional data. However, there are also studies in support of the BFLPE based on longitudinal data where the temporal ordering is more clear-cut and for true experimental studies in which class- or school-average achievement is experimentally manipulated.

Policy Implications

Our study greatly extends the generality of the negative effects of attending classes and schools where the average ability level of classmates is high, demonstrating the cross-cultural generalizability to primary school children as well as secondary school adolescents. These results also greatly expand the scope of support for

the universality of the BFLPE as a panhuman theory that has previously been based primarily on PISA data (Seaton et al., 2009). Indeed, our results suggest that the negative effects of school-average achievement based on PISA might substantially underestimate the results for more proximally relevant measures of class-average achievement based on TIMSS. Although theoretically important, these findings are worrisome, as ASC is well known to be an important predictor of academic choice and long-term engagement (Guay, Marsh, & Boivin, 2003; Marsh, 1991; Marsh & Craven, 2006; Marsh & O'Mara, 2010; Marsh & Yeung, 1997). Particularly when so many parents, teachers, and policy analysts uncritically assume that academic selective schools must automatically benefit the students who attend them, it is important to provide an alternative perspective based on strong theory and rigorous research. More generally, BFLPE research provides an alternative, contradictory perspective to educational policy on the placement of students in special education settings, which is a hotly debated topic in many countries throughout the world.

References

Abu-Hilal, M. M. (2001). Correlates of achievement in the United Arab Emirates: A sociocultural study. In D. M. McInerney & S. Van Etten (Eds.), *Research on sociocultural influences on motivation and learning* (Vol. 1, pp. 205–230). Greenwich, CT: Information Age.

Abu-Hilal, M. M., & Aal-Hussain, A. A. (1997). Dimensionality and hierarchy of the SDQ in a non-Western milieu: A test of self-concept invariance across gender. *Journal of Cross-Cultural Psychology, 28,* 535–553. doi:10.1177/0022022197285002

Abu-Hilal, M. M., & Bahri, T. M. (2000). Self-concept: The generalizability of research on the SDQ, Marsh/Shavelson model and I/E reference model to United Arab Emirates students. *Social Behavior and Personality, 28,* 309–322. doi:10.2224/sbp.2000.28.4.309

Alicke, M. D., Zell, E., & Bloom, D. L. (2010). Mere categorization and the frog-pond effect. *Psychological Science, 21,* 174–177. doi:10.1177/0956797609357718

Alwin, D. F., & Otto, L. B. (1977). High school context effects on aspirations. *Sociology of Education, 50,* 259–273. doi:10.2307/2112499

American Institutes for Research. (2005). *Reassessing U.S. international mathematics performance: New findings from the 2003 TIMSS and PISA*. Washington, DC: Author. Retrieved from http://www.air.org/files/TIMSS_PISA_math_study1.pdf

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science, 1,* 164–180. doi:10.1111/j.1745-6916.2006.00011.x

Bates, E. (1990). Language about me and you: Pronominal reference and the emerging concept of self. In D. Cicchetti & M. Beeghly (Eds.), *The self in transition: Infancy to childhood* (pp. 165–182). Chicago, IL: University of Chicago Press.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13,* 186–203. doi:10.1207/s15328007sem1302_2

Bornholt, L. J. (1997). Aspects of self knowledge about activities with young children. *Every Child, 3,* 15–18.

Bouffard, T., Markovits, H., Vezeau, C., Boisvert, M., & Dumas, C. (1998). The relation between accuracy of self-perception and cognitive development. *British Journal of Educational Psychology, 68,* 321–330. doi:10.1111/j.2044-8279.1998.tb01294.x

Bruner, J. (1996). A narrative model of self construction. *Psyke & Logos, 17,* 154–170.

Chapman, J. W., & Tunmer, W. E. (1995). Development of children's reading self-concepts: An examination of emerging subcomponents and their relation with reading achievement. *Journal of Educational Psychology, 87,* 154–167. doi:10.1037/0022-0663.87.1.154

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14,* 464–504. doi:10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233–255. doi:10.1207/S15328007SEM0902_5

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Damon, W., & Hart, D. (1988). *Self-understanding in childhood and adolescence*. New York, NY: Cambridge University Press.

Davis, J. (1966). The campus as a frog pond: An application of the theory of relative deprivation to career decisions of college men. *American Journal of Sociology, 72,* 17–31. doi:10.1086/224257

Davis-Kean, P. E., & Sandler, H. M. (2001). A meta-analysis of measures of self-esteem for young children: A framework for future measures. *Child Development, 72,* 887–906. doi:10.1111/1467-8624.00322

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist, 55,* 34–43. doi:10.1037/0003-066X.55.1.34

Diener, E., & Fujita, F. (1997). Social comparison and subjective well-being. In B. P. Buunk & F. X. Gibbons (Eds.), *Health, coping, and well-being: Perspectives from social comparison theory* (pp. 329–358). Mahwah, NJ: Erlbaum.

DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling, 9,* 327–346. doi:10.1207/S15328007SEM0903_2

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology, 47,* 309–326. doi:10.1111/j.2044-8317.1994.tb01039.x

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.

Eccles, J. S. (with Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C.). (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation: Psychological and sociological approaches* (pp. 75–146). San Francisco, CA: Freeman.

Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development, 64,* 830–847. doi:10.2307/1131221

Eder, R. A., & Mangelsdorf, S. C. (1997). The emotional basis of early personality development: Implications for the emergent self-concept. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 209–240). San Diego, CA: Academic Press. doi:10.1016/B978-012134645-4/50010-X

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12,* 121–138. doi:10.1037/1082-989X.12.2.121

Ertl, H. (2006). Educational standards and the changing discourse on education: The reception and consequences of the PISA study in Germany. *Oxford Review of Education, 32,* 619–634. doi:10.1080/03054980600976320

Fantuzzo, J. W., McDermott, P. A., Manz, P. H., Hampton, V. R., & Burdick, N. A. (1996). The pictorial scale of perceived competence and social acceptance: Does it work with low-income urban children? *Child Development, 67,* 1071–1084. doi:10.2307/1131880

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7,* 117–140. doi:10.1177/001872675400700202

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review, 87,* 477–531. doi:10.1037/0033-295X.87.6.477

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology, 95,* 124–136. doi:10.1037/0022-0663.95.1.124

Harter, S. (1983). Developmental perspectives on the self-system. In P. H. Mussen (Ed.), *Handbook of child psychology* (Vol. 4, 4th ed., pp. 275–385). New York, NY: Wiley.

Harter, S. (1998). The development of self-representations. In W. Damon (Ed.) & S. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 553–617). New York, NY: Wiley.

Harter, S. (1999). *The construction of the self: A developmental perspective*. New York, NY: Guilford Press.

Harter, S. (2006). The self. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 505–570). Hoboken, NJ: Wiley.

Harter, S. (2012). *The construction of the self: Developmental and sociocultural foundations* (2nd ed.). New York, NY: Guilford Press.

Hattie, J. (1992). *Self-concept*. Hillsdale, NJ: Erlbaum.

Hattie, J., & Marsh, H. W. (1996). Future directions in self-concept research. In B. A. Bracken (Ed.), *Handbook of self-concept* (pp. 421–462). New York, NY: Wiley.

Helson, H. (1964). *Adaptation-level theory*. New York, NY: Harper & Row.

Hopmann, S., Brinek, G., & Retzl, M. (Eds.). (2007). *PISA According to PISA*. Vienna, Austria: Verlag.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. doi:10.1080/10705519909540118

Huguet, P., Dumas, F., Marsh, H. W., Régner, I., Wheeler, L., Suls, J., . . . Nezlek, J. (2009). Clarifying the role of social comparison in the big-fish-little-pond effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology, 97,* 156–170. doi:10.1037/a0015558

Hutchison, G., & Schagen, I. (2007). Comparisons between PISA and TIMSS—Are we the man with two watches?In T. Loveless (Ed.), *Lessons learned: What international assessments tell us about math achievement* (pp. 227–261). Washington, DC: Brookings Institution.

Hyman, H. (1942). The psychology of subjective status. *Psychological Bulletin, 39,* 473–474.

James, W. (1963). *The principles of psychology*. New York, NY: Holt, Rinehart & Winston. (Original work published 1890).

Jerusalem, M. (1984). Reference group, learning environment and self-evaluations: A dynamic multi-level analysis with latent variables. In R. Schwarzer (Ed.), *Advances in psychology: Vol. 21. The self in anxiety, stress and depression* (pp. 61–73). Amsterdam, the Netherlands: North-Holland. doi:10.1016/S0166-4115(08)62115-9

Kagen, S. L., Moore, E., & Bredekamp, S. (Eds.). (1995). *Reconsidering children's early development and learning: Toward common views and vocabulary* (Report No. 95-03). Washington, DC: National Education Goals Panel.

Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research, 30,* 1–21. doi:10.1207/s15327906mbr3001_1

Lewis, M., & Brooks-Gunn, J. (1979). *Social cognition and the acquisition of self*. New York, NY: Plenum Press. doi:10.1007/978-1-4684-3566-5

Liem, G. A. D., Marsh, H. W., Martin, A. J., McInerney, D. M., & Yeung, A. A. (2013). The big-fish-little-pond effect and a national policy of within-school ability streaming: Alternative frames of reference. *American Educational Research Journal, 50,* 326–370. doi:10.3102/0002831212464511

Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2 × 2 taxonomy of multilevel latent contextual models: Accuracy–bias trade-offs in full and partial error correction models. *Psychological Methods, 16,* 444–467. doi:10.1037/a0024376

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods, 13,* 203–229. doi:10.1037/a0012869

Marsh, H. W. (1974). *Judgmental anchoring: Stimulus and response variables* (Unpublished doctoral dissertation). University of California, Los Angeles.

Marsh, H. W. (1984). Self-concept: The application of a frame of reference model to explain paradoxical results. *Australian Journal of Education, 28,* 165–181.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology, 79,* 280–295. doi:10.1037/0022-0663.79.3.280

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology, 81,* 417–430. doi:10.1037/0022-0663.81.3.417

Marsh, H. W. (1990). A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educational Psychology Review, 2,* 77–172. doi:10.1007/BF01322177

Marsh, H. W. (1991). Failure of high ability schools to deliver academic benefits commensurate with their students' ability levels. *American Educational Research Journal, 28,* 445–480. doi:10.3102/00028312028002445

Marsh, H. W. (2007). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology*. Leicester, England: British Psychological Society.

Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J. S., Abdelfattah, F., Leung, K. C., & Parker, P. (2013). Factorial, convergent, and discriminant validity of TIMSS math and science motivation measures: A comparison of Arab and Anglo-Saxon countries. *Journal of Educational Psychology, 105,* 108–128. doi:10.1037/a0029907

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391–410. doi:10.1037/0033-2909.103.3.391

Marsh, H. W., Chessor, D., Craven, R. G., & Roche, L. (1995). The effects of gifted and talented programs on academic self-concept: The big fish strikes again. *American Educational Research Journal, 32,* 285–319. doi:10.3102/00028312032002285

Marsh, H. W., & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 131–198). Orlando, FL: Academic Press.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1,* 133–163. doi:10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., Craven, R. G., & Debus, R. (1998). Structure, stability, and development of young children's self-concepts: A multicohort–multioccasion study. *Child Development, 69,* 1030–1053. doi:10.1111/j.1467-8624.1998.tb06159.x

Marsh, H. W., Debus, R., & Bornholt, L. (2005). Validating young children's self-concept responses: Methodological ways and means to understand their responses. In D. M. Teti (Ed.), *Handbook of research methods in developmental science* (pp. 138–160). Oxford, England: Blackwell. doi:10.1002/9780470756676.ch8

Marsh, H. W., Ellis, L., & Craven, R. G. (2002). How do preschool children feel about themselves? Unraveling measurement and multidimensional self-concept structure. *Developmental Psychology, 38,* 376–393. doi:10.1037/0012-1649.38.3.376

Marsh, H. W., & Hau, K.-T. (2003). Big-fish-little-pond effect on academic self-concept: A cross-cultural (26-country) test of the negative

effects of academically selective schools. *American Psychologist, 58,* 364–376. doi:10.1037/0003-066X.58.5.364

Marsh, H. W., & Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal–external frame of reference predictions across 26 countries. *Journal of Educational Psychology, 96,* 56–67. doi:10.1037/0022-0663.96.1.56

Marsh, H. W., Hau, K.-T., & Craven, R. G. (2004). The big-fish-little-pond effect stands up to scrutiny. *American Psychologist, 59,* 269–271. doi:10.1037/0003-066X.59.4.269

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 276–340). Mahwah, NJ: Erlbaum.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999). findings. *Structural Equation Modeling, 11,* 320–341. doi:10.1207/s15328007sem1103_2

Marsh, H. W., Kong, C.-K., & Hau, K.-T. (2000). Longitudinal multilevel models of the big-fish-little-pond effect on academic self-concept: Counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality and Social Psychology, 78,* 337–349. doi:10.1037/0022-3514.78.2.337

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research, 44,* 764–802. doi:10.1080/00273170903333665

Marsh, H. W., & O'Mara, A. (2010). Long-term total negative effects of school-average ability on diverse educational outcomes: Direct and indirect effects of the big-fish-little-pond effect. *Zeitschrift für Pädagogische Psychologie, 24,* 51–72. doi:10.1024/1010-0652.a000004

Marsh, H. W., & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology, 47,* 213–231. doi:10.1037/0022-3514.47.1.213

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish-little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review, 20,* 319–350. doi:10.1007/s10648-008-9075-6

Marsh, H. W., Trautwein, U., Lüdtke, O., Baumert, J., & Köller, O. (2007). Big-fish-little-pond effect: Persistent negative effects of selective high schools on self-concept after graduation. *American Educational Research Journal, 44,* 631–669. doi:10.3102/0002831207306728

Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology, 89,* 41–54. doi:10.1037/0022-0663.89.1.41

Marsh, H. W., & Yeung, A. S. (1999). The lability of psychological ratings: The chameleon effect in global self-esteem. *Personality and Social Psychology Bulletin, 25,* 49–64. doi:10.1177/0146167299025001005

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Routledge.

Möller, J., Streblow, L., & Pohlmann, B. (2009). Achievement and self-concept of students with learning disabilities. *Social Psychology of Education, 12,* 113–122. doi:10.1007/s11218-008-9065-z

Morse, S., & Gergen, K. J. (1970). Social comparison, self-consistency, and the concept of self. *Journal of Personality and Social Psychology, 16,* 148–156. doi:10.1037/h0029862

Muthén, B. O., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38,* 171–189. doi:10.1111/j.2044-8317.1985.tb00832.x

Muthén, L. K., & Muthén, B. O. (2013). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

Nagengast, B., & Marsh, H. W. (2012). Big fish in little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. *Journal of Educational Psychology, 104,* 1033–1053. doi:10.1037/a0027697

National Center for Education Statistics. (2008). *Comparing NAEP, TIMSS, and PISA in mathematics and science*. Retrieved from http://nces.ed.gov/timss/pdf/naep_timss_pisa_comp.pdf

Neidorf, T. S., Binkley, M., Gattis, K., & Nohara, D. (2006). *Comparing mathematics content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 assessments: Technical report* (NCES 2006-029). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Nicholls, J. G. (1979). Development of perceptions of own attainment and causal attributions of success and failure in reading. *Journal of Educational Psychology, 71,* 94–99. doi:10.1037/0022-0663.71.1.94

Olson, J. F., Martin, M. O., & Mullis, I. V. S. (Eds.). (2008). *TIMSS 2007 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Parducci, A. (1995). *Happiness, pleasure, and judgment: The contextual theory and its applications*. Mahwah, NJ: Erlbaum.

Parker, P. D., Schoon, I., Tsai, Y.-M., Nagy, G., Trautwein, U., & Eccles, J. S. (2012). Achievement, agency, gender, and socioeconomic background as predictors of postschool choices: A multicontext study. *Developmental Psychology, 48,* 1629–1642. doi:10.1037/a0029167

Penn, C. S., Burnett, P. C., & Patton, W. (2001). The impact of attributional feedback on the self-concept of children aged four to six years in preschool. *Australian Journal of Guidance and Counselling, 9,* 21–34.

Raykov, T., & Marcoulides, G. A. (2004). Using the delta method for approximate interval estimation of parameter functions in SEM. *Structural Equation Modeling, 11,* 621–637. doi:10.1207/s15328007sem1104_7

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley. doi:10.1002/9780470316696

Ruble, D. N., & Dweck, C. S. (1995). Self-conceptions, person conceptions, and their development. In N. Eisenberg (Ed.), *Review of personality and social psychology: Vol. 15. Social development* (pp. 109–139). Thousand Oaks, CA: Sage.

Russell, L., Bornholt, L., & Ouvrier, R. (2002). Brief cognitive screening and self concepts for children with low intellectual functioning. *British Journal of Clinical Psychology, 41,* 93–104. doi:10.1348/014466502163831

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York, NY: Chapman and Hall/CRC. doi:10.1201/9781439821862

Schwartz, S. H., & Bilsky, W. (1990). Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications. *Journal of Personality and Social Psychology, 58,* 878–891. doi:10.1037/0022-3514.58.5.878

Seaton, M., & Marsh, H. W. (2013). Celebrating methodological-substantive synergy: Self-concept theory and methodological innovation. In D. McInerney, H. W. Marsh, R. G. Craven, & F. Guay (Eds.), *International advances in self research: Vol. 4. Theory driving research: New wave perspectives on self-processes and human development* (pp. 161–181). Greenwich, CT: Information Age Press.

Seaton, M., Marsh, H. W., & Craven, R. G. (2009). Earning its place as a pan-human theory: Universality of the big-fish-little-pond effect across 41 culturally and economically diverse countries. *Journal of Educational Psychology, 101,* 403–419. doi:10.1037/a0013838

Seaton, M., Marsh, H. W., & Craven, R. G. (2010). Big-fish-little-pond effect: Generalizability and moderation—Two sides of the same coin. *American Educational Research Journal, 47,* 390–433. doi:10.3102/0002831209350493

Seaton, M., Marsh, H. W., Dumas, F., Huguet, P., Monteil, J.-M., Régner, I., . . . Wheeler, L. (2008). In search of the big fish: Investigating the coexistence of the big-fish-little-pond effect with the positive effects of upward comparisons. *British Journal of Social Psychology, 47,* 73–103. doi:10.1348/014466607X202309

Segall, M. H., Lonner, W. J., & Berry, J. W. (1998). Cross-cultural psychology as a scholarly discipline: On the flowering of culture in behavioural research. *American Psychologist, 53,* 1101–1110. doi:10.1037/0003-066X.53.10.1101

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55,* 5–14. doi:10.1037/0003-066X.55.1.5

Skaalvik, E. M., & Hagtvet, K. A. (1990). Academic achievement and self-concept: An analysis of causal predominance in a developmental perspective. *Journal of Personality and Social Psychology, 58,* 292–307. doi:10.1037/0022-3514.58.2.292

Stipek, D., & Mac Iver, D. (1989). Developmental change in children's assessment of intellectual competence. *Child Development, 60,* 521–538. doi:10.2307/1130719

Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., & Williams, R. M. (1949). *The American soldier: Adjustments during army life* (Vol. 1). Princeton, NJ: Princeton University Press.

Tymms, P. (2001). A test of the big fish in a little pond hypothesis: An investigation into the feelings of seven-year-old pupils in school. *School Effectiveness and School Improvement, 12,* 161–181. doi:10.1076/sesi.12.2.161.3452

Upshaw, H. S. (1969). The Personal Reference Scale: An approach to social judgment. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 4, pp. 315–370). New York, NY: Academic Press. doi:10.1016/S0065-2601(08)60081-7

Wedell, D. H., & Parducci, A. (2000). Social comparison: Lessons from basic research on judgment. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 223–252). Dordrecht, the Netherlands: Kluwer Academic. doi:10.1007/978-1-4615-4237-7_12

Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review, 12,* 265–310. doi:10.1016/0273-2297(92)90011-P

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., Freedman-Doan, C., & Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology, 89,* 451–469. doi:10.1037/0022-0663.89.3.451

Wu, M. (2009). A comparison of PISA and TIMSS 2003 achievement results in mathematics. *Prospects, 39,* 33–46. doi:10.1007/s11125-009-9109-y

Zell, E., & Alicke, M. D. (2009). Contextual neglect, self-evaluation, and the frog-pond effect. *Journal of Personality and Social Psychology, 97,* 467–482. doi:10.1037/a0015453

Received July 13, 2013
Revision received March 3, 2014
Accepted May 8, 2014 ■